DOCUMENT RESUME

ED 150 976 IR 005 660

AUTHOR Knapp, John; And Others
TITLE Establishment of Library Network On-Line Access to the Library of Congress MARC Data Base.
INSTITUTION Research Libraries Group, Branford, Conn.
REPORT NO BLG/SYS-75/0003
PUB DATE 14 Jan 76
NOTE 47p.

EDRS PRICE MF-$0.83 HC-$2.06 Plus Postage.
DESCRIPTORS *Cooperative Planning; Delivery Systems; *Library Networks; *On Line Systems; Pilot Projects; *Program Proposals; *Telecommunication

ABSTRACT

The Research Libraries Group proposes to establish a cooperative development program with the Library of Congress to provide computer-based, on-line access by library networks to the machine readable MARC file at the Library of Congress and data bases elsewhere. This document provides the proposal for Phase One, which would establish a communications link between members of the Research Libraries Group and the Library of Congress to test currently available data base access methods and provide empirical data for planning later phases. This phase is presented in terms of development, including economic considerations, benefits and technical approach; products; the project plan; personnel; and budget. Phases Two and Three are briefly described for better understanding of the role of Phase One in the total project. (Author/KP)

**********
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
**********

The Research Libraries Group

ED150976

# ESTABLISHMENT OF LIBRARY NETWORK ON-LINE ACCESS TO THE LIBRARY OF CONGRESS MARC DATA BASE

A Proposal for a Joint Pilot Project
conducted by the
Research Libraries Group
and the
Library of Congress

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Research Libraries group, Inc.

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

January 14, 1976

I 200 5-660



COLUMBIA UNIVERSITY LIBRARY · HARVARD UNIVERSITY LIBRARY
THE NEW YORK PUBLIC LIBRARY · YALE UNIVERSITY LIBRARY

# ESTABLISHMENT OF LIBRARY NETWORK ON-LINE ACCESS TO THE LIBRARY OF CONGRESS MARC DATA BASE

## PROJECT DESCRIPTION

- Contents -


| | | Page |
|---|---|---|
| 1. | SUMMARY | 2 |
| 2. | BACKGROUND | 4 |
| | 2.1 The Research Libraries Group | |
| | 2.2 The Library of Congress | |
| 3. | PHASE ONE DEVELOPMENT | 8 |
| | 3.1 General Description | |
| | 3.2 Economic Considerations | |
| | 3.3 Summary of Benefits | |
| | 3.4 Technical Approach | |
| 4. | PHASE TWO AND PHASE THREE DEVELOPMENT | 30 |
| | 4.1 General Description | |
| | 4.2 Summary of Features | |
| 5. | SUMMARY OF PHASE ONE PROJECT PRODUCTS | 35 |
| | 5.1 Computer Software | |
| | 5.2 User Documentation | |
| | 5.3 Reports and Publications | |
| 6. | PHASE ONE PROJECT PLAN | 37 |
| 7. | PHASE ONE PROJECT PERSONNEL | 41 |
| | PROJECT BUDGET | Follows page 41 |

APPENDICES

A. The RLG Bibliographic System: Design Requirements
B. The Library of Congress as the National Bibliographic Center
C. A Model for Computing Delays in Obtaining Service with a Finite Number of Point-to-Point Lines
D. User's Guide to Constructing Author/Title (3,3) and Title (3,1,1,1) Search Keys
E. User's Guide to On-line Search of MARC Data
F. Phase One Software Design-Logic
G. The MARCARD Utility Program
H. Development of a One Year Budget Figure for Use of the Library of Congress Computer
I. Yale's Use of OCLC through NELINET Using OCLC Announced Prices for 1976/1977

## 1. SUMMARY

The Research Libraries Group proposes to establish a cooperative development program with the Library of Congress, the ultimate objective of which is to provide computer-based, on-line access by library networks to the machine-readable files at the Library of Congress and elsewhere. These bibliographic records will be used in support of library collection development, acquisitions, cataloging, and interlibrary loan procedures. Substantial savings and increased effectiveness over alternative modes of access are anticipated in a configuration where the network will store and provide access to only those records which relate to the holdings of its member institutions. Automatic, computer-based linkage to more comprehensive files at the Library of Congress and in other library networks will provide bibliographic data for items not yet held by the local networks. The scholarly and research community will be the eventual beneficiary of these activities as the service advantages of computer-based library catalogs begin to accrue in the form of prompt availability of current literature, more powerful tools for bibliographic search, and direct bibliographic access to the shared resources of several libraries.

Phase One of the program will establish a simple communications link between designated members of the Research Libraries Group and the Library of Congress to test currently available data base access methods and to provide empirical data for planning later project phases. In addition, on-line record update and off-line catalog card production components will be installed to support the card catalogs of the participating libraries and, at the same time, to provide new data for the New York Public Library book catalog. Phase One has potential to provide an effective cataloging service for the Research Libraries Group members, and it is intended that the service be continued on a cost recovery basis after the expiration of the pilot project.

Phases Two and Three of the program will lead to installation of the computer hardware and software components at LC and RLG needed to achieve communications links not only with the Library of Congress but also with other data bases of interest to research libraries, i.e., other library networks and the abstracting and indexing services. This capability will permit bi-directional data exchange which will allow the Research Libraries Group to make available its cataloging records to other libraries. Standard protocols and software modules will be defined to isolate communications traffic from possible incompatibilities between terminals, computers, operating systems, and programming languages. To assure general application of the system for other organizations, it is anticipated that other library networks will be invited to join the Research Libraries Group and the Library of Congress in developing the design specifications for this communications facility.

This proposal seeks funds for Phase One development only. However the rationale for Phase One is best understood within the context of the total program and, therefore, a general description of Phases Two and Three is included in section 4. Support for Phase Two will be sought separately but concurrently with Phase One, and these two development phases can be conducted virtually in parallel depending on availability of funding.

Phase One is scheduled to operate for 18 calendar months on a budget of $394,448.

## 2. BACKGROUND

### 2.1 The Research Libraries Group

The Research Libraries Group is a consortium comprising the Libraries of Columbia, Harvard, and Yale Universities and the Research Libraries of New York Public Library. It was established out of the recognition that neither significant increases in library purchasing power nor reductions in demand for library services are probable in the reasonable future and that an alternative to the heretofore dominant philosophy of institutional self-sufficiency needed to emerge.

To this end, RLG has launched programs in support of shared access on a reciprocal basis to the combined collections of the four institutions--an inventory of approximately 26.5 million volumes. In addition, RLG is actively planning for future expansion of these resources in a more rational way by avoiding unnecessary duplication among the four libraries. Because periodicals require over half the funds budgeted for acquisitions and a substantial portion of the processing budget for the ancillary costs of receiving, claiming, cataloging, binding, storage, etc., special attention is being given to existing duplication and future recommendations for new subscriptions. Also, expensive items costing more than a specified amount are being jointly reviewed to reduce duplication. Since research libraries have responsibilities not only to provide access to their collections but also to assure continued access by providing appropriate physical protection, therefore integral to RLG programs is coordinated planning for preservation of its members' materials.

Shared access, coordinated collection development, and preservation programs need the support of a well developed system of bibliographic control which is the product of the cataloging activities in the libraries. Such a system has the dual objectives of enhancing service to RLG library users by making known the existence and location of materials within the library and in other member institutions, and of reducing the cost of cataloging by taking bibliographic data available from a central and authoritative source and applying it to the local cataloging process.

RLG intends to develop a computer-based system for the processing and control of bibliographic information. Such a system will allow RLG to make effective use of machine-readable cataloging available from the Library of Congress and of other cataloging generated within the RLG membership. Initial requirements for this system have been described and are attached as Appendix A. These requirements imply an endeavor which is the most ambitious yet undertaken by any consortium. The ultimate goal is to have a single authoritative data base of holdings of all four institutions available on-line. This on-line union catalog will be used as the central control mechanism for the various shared resources programs and is viewed as enhancing service to library users and as providing an alternative to the costly and labor-intensive maintenance of local card catalogs. It is hoped that many of these requirements can be met by transferring and adapting components of computer-based systems which have been developed independently by RLG members or elsewhere.

Fundamental to an RLG bibliographic system is standardization based on national and international standards and codes as reflected in Library of Congress cataloging practice and in the MARC II format for machine-readable records. Already three RLG members adhere to Library of Congress practice with only minor variations easily accommodated in a common system and the fourth member, Harvard University, has recently announced that its central research collection in the Harvard College Library will be adopting LC practice in FY 1976/77. It is understood among RLG members that RLG standards will be consistent with national standards and that where local practice varies, members must accommodate both RLG and local requirements.

While the Phase One proposal is focused on a program to identify and transfer records from a national data base at the Library of Congress to the Research Libraries Group, this development should be viewed as the first step toward the creation of a comprehensive computer-based bibliographic processing system which will provide a wide variety of services for present and future members of RLG.

Placed within the context of RLG's long-range plans for mechanization, Phase One itself has inherent possibilities for being subsequently expanded in areas where a mutuality of interest should exist between RLG and the Library of Congress. It has been estimated that RLG libraries give original cataloging to over 100,000 titles a year, titles for which it has been determined that cataloging in most instances will not be readily available from the Library of Congress. Having committed itself to following the same bibliographic standards as LC, RLG would like to develop the capability to use a common machine-readable authority file with the Library of Congress, thus making it possible for our unique cataloging records to be integrated into the national data base and made available to all libraries. As an extension of this concept, it might be possible for LC and RLG to reach an agreement under which RLG would be given national responsibility to provide cataloging information for selected categories of material.

In addition, the Library of Congress has developed a computer-based system to locate and control recently acquired material which is in the cataloging process; the system provides access to and maintenance of the LC Process Information File. On-line access to this file by libraries outside the Library of Congress would substantially increase the number of titles for which at least partial bibliographic information could be made available soon after publication.

Finally, experience in the use of other bibliographic data bases indicates that their utility for selection and pre-order searching in acquisitions work, and for locating titles wanted on interlibrary loan, is equal in importance to the cataloging function. The RLG program with the Library of Congress could be extended to provide these capabilities.

## 2.2 The Library of Congress

The Library of Congress, as the national library, has since 1901 been the primary creator and distributor of cataloging data on printed 3x5 cards. This role has been considerably enhanced by its National Program for Acquisitions and Cataloging which has provided bibliographic descriptions for an increasingly larger portion of the cataloging done by libraries in the United States. In 1969 a further significant step forward was made with the introduction of the MARC* Distribution Service, i.e. machine-readable cataloging distributed weekly on magnetic tape. The existence of this data has made possible many new and innovative services by stimulating automated systems in American libraries. It has also created a new environment for library cooperation which bypasses many of the problems of the past.

The distribution of bibliographic data in machine-readable form has resulted in developments which cannot be thought of simply as more efficient extrapolations of the Library's traditional card distribution service. By virtue of the flexibility permitted by automated systems and of the immediacy of on-line systems, a whole new environment has been created for library development. Cooperative efforts between libraries, which previously would have required centralization of operations, are now feasible with decentralized systems. This has led to possibilities for resource sharing hitherto unattainable, both in terms of labor in technical processing centers, and what may perhaps prove even more significant, of material resources via shared collection development and improved access for library users.

In response to this new environment the Library of Congress has launched a planning effort to enhance its traditional modes of cataloging distribution with services tailored to the needs of libraries participating in computer-based networks. The current concept for a "National Bibliographic Service" would include first the maintenance of a large data base of Library of Congress cataloging records converted to machine-readable form both by the Library of Congress (MARC) and by other libraries. Later the service could include cataloging originated by other libraries, both national and international, and authority records. An amplification of this concept can be found in the 1975 statement by William J. Welsh, Director of the Precessing Department, Library of Congress, which is included as Appendix B. Initially this LC data base would be used for expanded coverage of the current magnetic tape distribution services. More important would be the provision of remote on-line access to this data base allowing two-way transmission of records.

Consistent with this goal, the Library of Congress has in the past few years devoted considerable staff effort to building on-line capabilities. Recently the Library has installed for its own use a system which allows on-line search of the MARC data base using author and title, or card number, as the means of access. This facility is now used sucessfully throughout the Library. The development outlined in this proposal would test the feasibility of applying this internal capability to the National Bibliographic Service.

---

* Acronym for machine-readable cataloging

It is significant that both the statement concerning a National Bibliographic Service at the Library of Congress (Appendix B) and the recommendations published by the National Commission on Libraries and Information Services consider network development as one of the essential elements in the improvement of information services in this country. The utility of the present proposal is not limited to the Library of Congress and the Research Libraries Group, but has the potential for serving as a model which can be later extended to any other library network which wants to obtain bibliographic records from the national library cataloging data base.

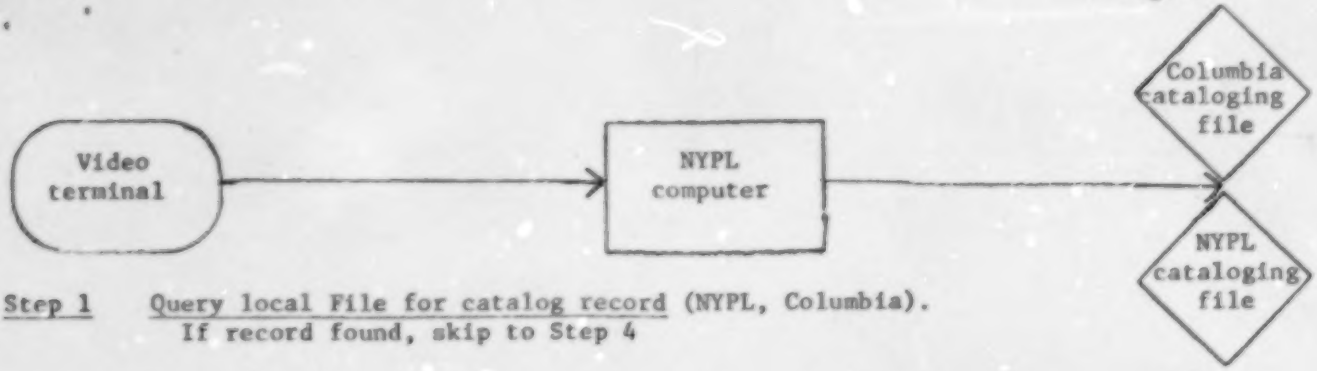
## 3. PHASE ONE DEVELOPMENT

### 3.1 General Description

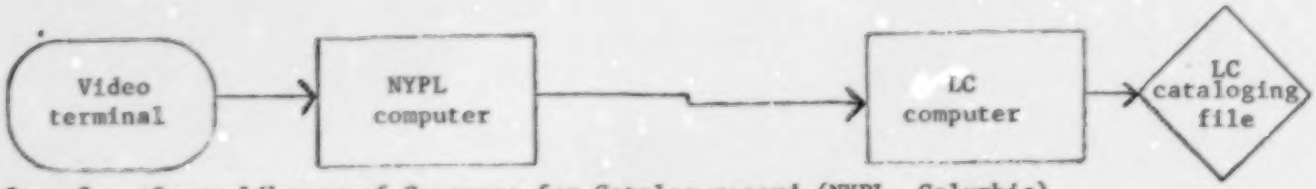
As part of the cataloging process for new acquisitions, libraries routinely attempt to determine whether an item has been cataloged by the Library of Congress; if it has, they use that cataloging record as source data for integrating a new item into their collection. There exist obvious economies in not having to give every item original cataloging. Against these economies, however, libraries must balance the cost of locally maintaining large files of Library of Congress cataloging data in card and book form and laboriously searching these files to find the record required. An alternative form of Library of Congress data has been developed in the past six years as the Library of Congress has been converting an increasing portion of its cataloging data to machine-readable form and distributing this information weekly to subscribing libraries and library networks. Starting only with English language titles in 1969, the Library of Congress MARC distribution is expected to include cataloging for all Roman alphabet languages by 1980. This distribution service still requires local maintenance of a large file of source records (now a computer-based file) of which only approximately 30-50 percent will be used.

The concept of Phase One is quite simple. The proposition to be tested is that it is more economic to access the Library of Congress computer-based files on-line in Washington, D.C. than it is to maintain these large files locally. Conversely, from the point of view of the Library of Congress, the project will investigate the economies of providing such a service to the nation's libraries through library network facilities. These economic implications are discussed more fully in section 3.2 below.

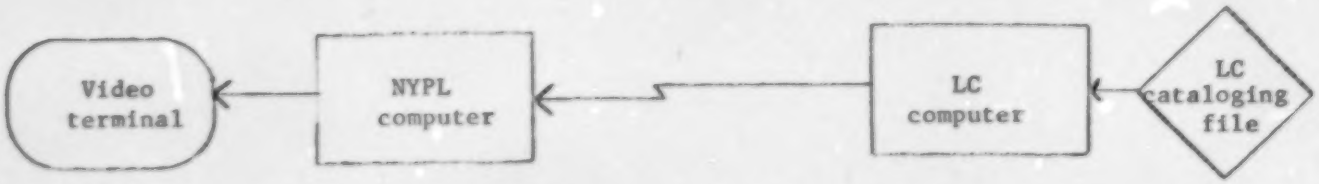
Specifically, under the auspices of the Research Libraries Group, video and printer terminals will be installed at the Columbia University Libraries and at the New York Public Library. These terminals will be connected to a computer currently maintained by the New York Public Library. A communications link will be established between this computer and the computer at the Library of Congress. Using this facility the two libraries can negotiate searches for cataloging records first on the local data base and, if unsuccessful, on the Library of Congress data base located in Washington, D.C. (See general procedure diagram on the next page.) When the proper cataloging record is found, the full record will be transmitted to the computer in New York where it may be immediately modified to suit local requirements. Some terminals will have printer devices attached to them so that data displayed on the terminal screen can be printed when required. This "hard copy" version of the record will facilitate consultation with library manual files away from the terminal site which may be necessary to make proper modifications to the record. When complete, the record will be transferred to separate cataloging files maintained by the two libraries from which catalog cards or book catalogs will be produced. At the appropriate time during the project the other two members of RLG, Harvard and Yale, will be linked to the facility if experience shows they could benefit from its services.



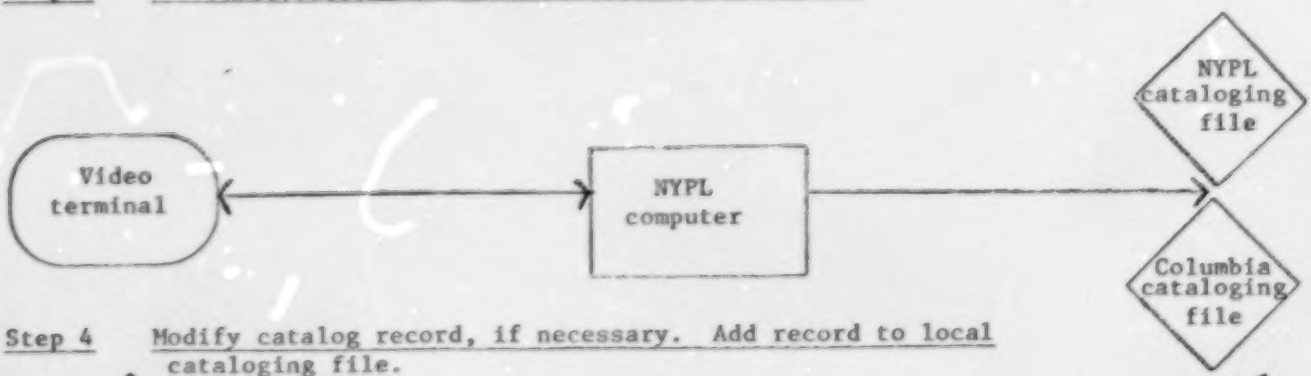

Step 1 Query local File for catalog record (NYPL, Columbia).
If record found, skip to Step 4



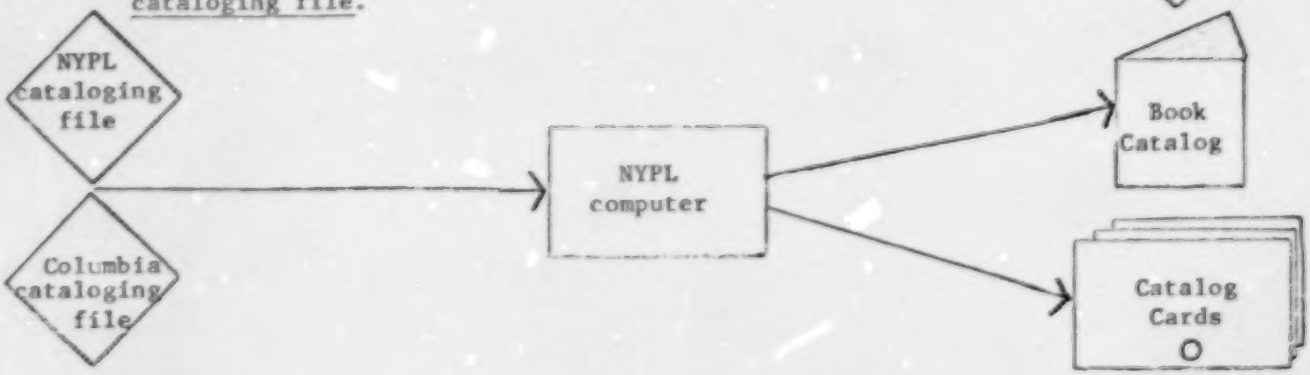

Step 2 Query Library of Congress for Catalog record (NYPL, Columbia).
If record not found, revert to current cataloging procedures. These may include searching for the item again on-line after a suitable wait.




Step 3 Transmit catalog record from Library of Congress




Step 4 Modify catalog record, if necessary. Add record to local cataloging file.




Step 5 Produce local cataloging products.

This pilot project has the advantages of being relatively inexpensive and requiring only a short time to implement. Little additional equipment (other than modems and line control adaptors for existing transmission control units, and terminals, printers, and modems in the libraries) is required, and only limited additional software will be needed. The project uses components of existing hardware and software developed separately by RLG members and integrates them with minimal development effort. These components are: update of bibliographic records on-line using a video terminal, a procedure developed by New York Public Library; and computer printing of 3x5 catalog cards according to very flexible controls which allow output tailored to specific institutional requirements, a procedure developed by Columbia University Libraries. Furthermore, no new standards or modifications to existing standards are to be promulgated by this project phase; rather the opposite is suggested. Where relevant standards exist they will be adopted in order to provide a basis for the analysis of current standards to be undertaken in project Phase Two.

The work to be completed under Phase One is viewed as the initial development of an RLG system. It will significantly impact the cataloging processes in the participating libraries, permitting them to access Library of Congress data, to modify it to suit the book-in-hand (e.g., a different edition) and local needs (e.g., local call number), and to produce cataloging products. It will allow RLG to begin building separate cataloging data bases of its members' holdings which later can be integrated into one file consistent with the ultimate development objective of building a single, coherent on-line catalog for RLG. A card production service for maintenance of local card catalogs is a very practical but intermediate objective only.

### 3.2 Economic Considerations

This project will focus on two cost/benefit issues important to one of RLG's long-range goals; lowering the cost of cataloging in research libraries. These are: 1) the optimal mix of computing, storage, and communications costs and their influence on data base configuration, and 2) the effect on cataloging costs of on-line access to Library of Congress MARC data.

#### 3.2.1 Potential Facility Configuration

Three systems designs present themselves as potential RLG facility configurations, their distinction lying in the location of the LC source cataloging records and the associated indexes. These indexes and their use need to be described in order to fully understand the system alternatives.

Broadly speaking there are two ways for a user to search a cataloging data base: a precise search by LC Card Number (which is usually printed on the verso of the title page of books from domestic trade publishers)

and a less precise search based on the author or title of the work. The index search based on the LC Card Number will produce a definitive yes-or-no response from the system: the item is either present in the data base or it is not. Author/title index searches often require several transactions between the user and the system in order to determine whether the desired record is or is not in the data base. This process of negotiation is required because the search key entered is usually based on some combination of the initial letters from the author and title words. Similar "pre-coordinated" search keys are stored in the system as an index to the data base and often these few letters match more than one record in a large data base. System designers effect a compromise between the number of characters that need to be keyed by the user and the number of times the resulting key will match on more than one record. When a multiple match occurs, the system will display enough pertinent information for each record so the user can choose a single record or reject them all. This choice can require several transactions to resolve the search.

The combination of multiple search requirements and the need for an expedient means of changing and adding to the data base is best served by establishing the full text of the records in a file in an arbitrary order and then building several ordered index files with pointers to the full records. Such indexes can then give access by a variety of methods: author, title or LC Card Number as proposed in this project; possibly also subject, series, or International Standard Book Number. As noted above, when using an index to resolve the search it is often necessary to display more bibliographic information than is found in the original search key. For system efficiency it may be best to store this information along with the search key in the index, but such system considerations can result in indexes which themselves require 50% of the storage space of the original full records. The cost of storing and maintaining the file of full records and the index files (i.e., the data base of bibliographic information) is a major cost element in any on-line system. It is the location of, and means of providing access to, the LC cataloging source data which is of interest in the present description of three potential configurations.

In each of the alternatives described below it is assumed that records for titles held by RLG member libraries are maintained locally with a local index to them. Our inquiry is directed toward access to records for materials not yet held by RLG libraries but already cataloged by the Library of Congress. The alternatives are:

1. Source Records and Index at LC (the proposed project);

2. Source Records at LC, Index Local;

3. Source Records and Index Local.

In the first alternative the local data base is first searched via the local index. If the desired record is not found locally, the search is re-entered and routed to LC. There is no cost of maintaining unused LC MARC data locally, but there are costs associated with communication to LC and with the computer and user having to reprocess the search.

In the second alternative, a combined index both to local records and to MARC records located at LC would require only a single search. If the record exists locally, the user will obtain it directly; if instead the index points to a record at LC, this record will be obtained from LC. The communication costs in this alternative will be lower because the search will be negotiated and resolved locally and only a single transaction will be sent to and received from LC. However, there will be added storage costs of maintaining the LC MARC index integrally with the local index, and the index would require storage for sufficient bibliographic information to resolve a search without access to the full text of the record.

In the third alternative, a combined index and both the local and LC MARC records are maintained locally. There would be no communications costs, but there would be the storage costs of maintaining the MARC index and the unused LC MARC records.

The cost elements for each alternative are summarized below:

1. Costs of Maintaining Source Records and Index at LC. The proposed project will implement this first alternative; therefore, directly applicable data will be obtained for each of the cost items:

   a. communication lines and associated equipment (modems, etc.) adequate to handle the traffic with reasonable response time;

   b. LC charges for processing searches in its computer;

   c. labor required for repetition of user search negotiation when a record is not available locally.

Presumed savings over other alternatives:

   d. LC data base and index are not maintained locally.

2. Costs of Maintaining Source Records at LC, Index Locally. The proposed project will not directly implement this alternative, but its associated costs can be extrapolated from the data gathered in the operation of the first alternative. For example, the cost of maintaining the LC index locally can be developed from the cost of maintaining the index to local holdings. Similarly, the reduction in communications requirements can be derived from analysis of the traffic in the first alternative. The following costs will be accounted for:

   a. communications lines and associated equipment which will be less expensive because all requests for records at LC will be handled by a single transaction;

b. LC charges for processing, reduced for the same reason;

c. local computer processing required to load and periodically maintain the LC MARC portion of the index;

d. local computer memory and disk storage required to maintain the LC MARC portion of the index.

Presumed savings over other alternatives:

e. only a single search needs to be negotiated by the user;

f. fewer communications lines will be required.

3. Costs of Maintaining Source Records and Index Locally. As with the second alternative, though the proposed project will not directly implement this approach, the costs can be extrapolated from the data gathered in the operation of the first alternative. The following costs will be accounted for:

a. local computer processing required to load and periodically maintain the LC MARC data base and index;

b. local computer memory and disk storage required to maintain the LC MARC data base and index.

Presumed savings over other options:

c. no costs of communication to LC;

d. only a single search needs to be negotiated by the user;

e. no LC charges for processing.

### 3.2.2 Effect on Cataloging Costs

Careful analysis of the effect on cataloging costs of the service to be offered under this proposal by RLG to its members is one of the primary purposes of this project phase. It must be emphasized that the proposed service affects only a portion of the total cataloging in the research libraries served by RLG. An economic study covering the full range of cataloging activities within RLG has been started by the RLG staff. This study is fundamental to planning the RLG on-line union catalog because from it we can configure the scope of the system required (number of terminals, CPU power, storage space, transaction loads, etc.) and determine the amount of money now being spent against which we can target future savings. Such "savings" may be no more than to slow the rate of inflationary increase by making cataloging and catalog maintenance less labor intensive through the assistance of the computer, a technology whose costs have continued to decline. This study is not complete; therefore, we are not prepared to make any statements about the long range cost benefits of an RLG union catalog although we remain optimistic that these benefits will be demonstrated.

This optimism has been supported, however, by our analysis of this Phase One project which shows sufficient cost benefits to RLG libraries to justify the effort on its own even without reference to future development. As shown below, some RLG libraries will accrue dollar savings by using the system while others already using a similar service offered by the Ohio College Library Center can participate at no added cost. However, we must not lose sight of the most important benefit which will be to have all four libraries cataloging on a single system, sharing one another's data, and thereby launching a significant cooperative enterprise.

The following economic analysis shows: 1) a one year operational budget for the system, and from that what must be charged for each record used in order to get full cost recovery; and 2) a model which compares current manual cataloging systems with the proposed computer-assisted system and which shows related costs of each.

3.2.2.1 One Year Operational Budget for the RLG System

1. Projected Use of the System

| | | |
|---|---|---|
| New York Public Library | 31,200 | cataloging records |
| Columbia Library | 23,000 | " " |
| Yale University | 32,000 | " " |
| Harvard University | 32,000 | " " |
| | 118,200 | total records used |

The figure for NYPL is based on actual current use of LC MARC in conjunction with their automated book catalog system. Columbia's use is projected from an analysis of current acquisitions and cataloging activity. Yale's figure is based on their current use of the Ohio College Library Center system and increased to account for expanded coverage of LC MARC during 1976. Harvard's figure is based on use assumed to be similar to Yale's.

2. Budget for NYPL and Columbia - 54,200 records used

| | | | |
|---|---|---|---|
| 1. | Cost on LC System | $41,400 | See Appendix H for development of this figure. |
| 2. | Cost on RLG System | 51,750 | 1.25 x LC cost |
| 3. | Communications | 17,000 | |
| 4. | Terminal amortization | 18,667 | 3 years for 10 terminals, 4 printers |
| 5. | Terminal maintenance | 8,400 | $50/month x 14 |

(continued on next page)

| | | | |
|---|---|---|---|
| 6. | Salaries | | |
| | a. Operations | 12,000 | .5 senior programmer/ analyst |
| | b. RLG Administration | 14,200 | |
| 7. | Capital Budget | 5,420 | $.10 per record used |
| | | $168,837 | |

Note: Terminals are amortized at three years because terminal technology is developing rapidly and early replacement may be advisable. The RLG capital budget would be used for future enhancements to RLG systems. The amount per record would be set by the RLG Executive Committee as they think appropriate and is included here to suggest establishment of a principle.

To gain full cost recovery with this budget, RLG would have to charge $3.12 per record used plus the cost of the computer produced catalog cards (on average $.40 per set). This budget has used the current charge of 23¢ per transaction levied by the Library of Congress as the major determinant of the computer costs. Any change in that pricing policy could significantly affect the cost mix. The most speculative figure in this budget is item #2, RLG computer costs.

3. Budget for NYPL, Columbia, Yale and Harvard - 118,200 records used.

| | | | |
|---|---|---|---|
| 1. | Cost of LC System | $ 64,481 | See Appendix H |
| 2. | Cost of RLG System | 112,857 | Proportional to 2. above by number of records used |
| 3. | Communications | 21,000 | Added lines and modems for Yale and Harvard |
| 4. | Terminal amortization | 37,334 | 3 yrs for 20 terminals, 8 printers |
| 5. | Terminal maintenance | 16,800 | $50/month x 28 |
| 6. | Salaries | | |
| | a. Operations | 24,000 | |
| | b. RLG Administration | 14,200 | |
| 7. | Capital Budget | 11,820 | |
| | | $302,492 | |

To gain full cost recovery with this budget, RLG would have to charge $2.56 per record used plus the cost of computer produced catalog cards. Economies from full use of the system by four libraries are evident. Because an assumed 50% title duplication rate exists between the four, the use of the LC computer and the communications link to Washington, D.C. does not rise in direct proportion to the total records used in the system. Also management of the system for four libraries requires little extra effort than for two. Other economies of scale may be present and discovered in the course of operating the system (e.g. cost of the RLG system).

The charge of $2.56 per record used (or $2.96 including catalog cards) compares favorably with costs during fiscal year 76/77 for similar services offered by NELINET (New England Library Information Network) using the Ohio College Library Center system in Columbus, Ohio. Appendix I presents an analysis of the costs Yale University can expect during the next fiscal year which range from $2.43 to $2.59 per record depending on the number of records used. Pricing of the RLG services will present several important policy issues to the RLG Executive Committee, such as: to what extent must RLG remain price "competitive" with similar services in other networks, by what amount should RLG libraries pass on its "savings" to RLG for future network solvency and development, and to what extent should RLG libraries be expected to pay a short-range premium to subsidize long-range cooperative endeavors.

Pending resolution of such questions, we do plan to charge for services as soon as routine delivery begins during Phase One. However, these charges initially may be set to recover less than the full cost of the system in recognition of the high costs entailed by a library switching to a new and experimental system and of the fact that economies of scale will not have taken effect in the early, limited use of the system. Project funds will be used during the operational phase of the project to subsidize these costs. Monies received from RLG members during the project will be used for future development for the RLG bibliographic system.

3.2.2.2 Comparison of Manual and Computer-Assisted Systems

Cataloging in a research library is a complex process which involves many individuals with various levels of proficiency and salaries handling a bibliographic item before it finds its way to the shelves properly inventoried in the library's catalogs. We will not undertake to present a full description of this process here but will show only those activities which will be directly impacted by the computer-assisted system.

1. Pre-Cataloging Search. In a manual system the item is first searched both in the Library of Congress source cataloging files (card and book form) and in the local "official" catalog, the latter to determine if the item is an added copy or a new edition of a work already held. The purpose of this search is to determine whether the item will be cataloged with source copy available or will be given original cataloging. If copy is found it is placed in the book and passed on to cataloging for subsequent processing. With the computer-assisted system offered during Phase One, items

within the scope of LC MARC (i.e. within date and language boundaries) will be searched first on the terminals. When the proper bibliographic record is found, the record will be printed by a terminal printer, and this copy placed in the book and passed on for cataloging. Our model assumes that the selection of items for search on the terminal can be controlled so that 75% of the items searched will have matching records on the data base. Usually the search of the official catalog is deferred when using the computer-assisted method and is combined with the searching performed as part of the cataloging activity. This combination is possible because most libraries are able to organize the processing flow so that the searching and cataloging activities occur with little delay between the two steps; i.e. within 24 hours.

2. Depository Filing. As discussed before, research libraries maintain large manual files of Library of Congress cataloging which they use as source copy in support of their own acquisitions and cataloging. They receive approximately 235,000 to 250,000 catalog cards per year from the Library of Congress which have to be organized for use in local files. Even though the cards are received free under the National Program for Acquisitions and Cataloging, the filing of these cards represents a considerable expense which has to be amortized across the many fewer titles cataloged than cards filed. With the computer-assisted system, libraries will have the option to leave aside, unfiled, the approximately 144,000 cards containing cataloging which also appears on the LC MARC data base.

3. Cataloging with Copy. Bibliographic items for which source cataloging copy has been found are usually cataloged by an organizational unit separate from the original cataloging activity. Staff in this unit often are supervised by a professional librarian but are not themselves working at this rank. The source cataloging data is searched thoroughly against the official catalog in order to integrate the new item consistently into the collection as a whole. All catalog headings are checked, names, series, subjects, and the complete call number assigned against the shelf list. After appropriate modifications are made to the source data to suit the book-in-hand or to fit the record consistently into the local catalog, the cataloger prepares a master catalog card either by typing the whole card or modifying the LC printed card. In the computer-assisted method, the procedure would remain essentially the same, except that instead of preparing a master card, the cataloger would use a terminal to call up the record, make modifications on the video screen, and store the new record on the data base from which computer printed cards are produced.

4. Card Production. Most research libraries reproduce the master unit card on photocopy equipment making enough copies of the master to account for all catalog entries. After duplication the cards are processed by a typing unit which types the headings for each catalog entry at the top of the cards and performs any other tailoring of individual cards for particular catalog locations. By having the computer print the catalog cards, these activities are completely replaced.

5. Arranging for Filing. The card sets produced in the step above have to be batched and arranged first by the catalog into which they are to be filed and then by rough alphabetical order within each catalog. The computer card printing system can deliver the cards already arranged in this order thus eliminating this activity.

Columbia University Library currently has a completely manual cataloging operation which we have studied, developing unit costs per title for the activities described above. Yale University Library for the past year has made full use of a computer-assisted system using the facilities of the Ohio College Library Center. Therefore, we can apply the Yale experience to the Columbia model and estimate the effect of the computer assisted system as follows:

| | Current Manual | | Future Computer-assisted | |
|---|---|---|---|---|
| | Units/Hour | Cost/Title | Unit/Hour | Cost/Title |
| 1. Pre-cataloging search | 4.2 | $1.58 | 20.0 | $ .33 |
| 2. Depository filing | | .21 | | -- |
| 3. Cataloging with copy | 1.4 | 5.13 | 2.0 | 3.59 |
| 4. Card Production | | 1.25 | | .40 |
| 5. Arranging for Filing | | .19 | | -- |
| | | $8.36 | | $4.32 |
| Computer system charges | | -- | | 2.56 |
| | | $8.36 | | $6.88 |

According to this model the computer-assisted system should eventually save Columbia University approximately $1.48 per title -- not dramatic but amounting to $34,040 per year when applied to the expected 23,000 titles. Note that item 3 is probably the most important cost element affecting the savings potential. We must watch carefully to see if the overall productivity of the cataloging with copy unit can be increased from 1.4 titles per hour to 2.0 titles per hour by using the terminal system. Yale's experience with OCLC gives us confidence that savings indicated above have some validity. Projecting into fiscal 1976/77, Yale's use of the OCLC system at 32,000 records per year will follow the model on the following page.

| | Manual | | Computer-assisted | |
|---|---|---|---|---|
| | Units/Hour | Cost/Title | Units/Hour | Cost/Title |
| 1. Pre-cataloging search | 4.8 | $1.25 | 25 | .24 |
| 2. Depository filing | | .37 | | -- |
| 3. Cataloging with copy | 1.1 | 5.82 | 2.1 | 3.05 |
| 4. Card production | | .94 | | .40 |
| 5. Arranging for filing | | .26 | | -- |
| | | $8.64 | | $3.69 |
| Computer system charges (Fiscal 76/77) | | -- | | 2.43 |
| | | $8.64 | | $6.12 |

According to the experience at Yale, they will save approximately $2.52 per title during fiscal 76/77 by use of the computer-assisted system which when applied to the 32,000 titles they expect to process on the system next fiscal year is $80,640. Yale reports that an important added benefit has been to reduce the time a title spends in cataloging and processing from six weeks to two weeks.

Special attention should be paid to the effect the computer-assisted system has on the labor components of the costs. In the manual system almost all of the costs are for labor (except some materials costs in card production) while in the computer-assisted system labor comprises approximately 53 to 56% of the costs. As a long term benefit, this activity in the libraries will not continue to be as directly influenced by the 5-7% annual inflationary rate for staff salaries. At the same time all indications are that the cost performance of computers, storage equipment, and terminals will improve which may permit us to hold the line on computer system charges.

New York Public Library's cataloging operation does not fit the model presented above because it is already using an automated book catalog system, and therefore, determining the impact on costs of on-line access to LC MARC would require a separate study. Our guess is that the savings will be marginal in the cataloging operation, but we must not overlook the benefit to NYPL of income derived from RLG for sharing their computer with other libraries.

The Harvard College Library of the Harvard University Library system is currently a user of the OCLC system for a small portion of its cataloging. Some of its smaller libraries (e.g. the Lamont undergraduate collection) use Library of Congress cataloging standards, and titles for these libraries are processed by the computer-assisted system. However, as mentioned earlier,

Harvard College Library has announced its intention to adopt LC standards for its central research collections during fiscal year 1976/77. This means that efficient use of the computer-assisted system will be possible for a larger number of titles. Therefore, Harvard stands to show savings even in the operation of its manual system by use of LC data without regularly changing the classification number and subject headings. Also, full use of the computer-assisted system ought to accrue additional savings at least equal to those projected for Columbia. There are many other libraries at Harvard which could also benefit themselves and RLG by use of the computer-assisted system; however, their interest and the potential benefits to them have not been studied in preparation for this proposal. Contact will be maintained with these libraries, and their participation welcomed at any time that it is clear that the computer system can perform successfully for them.

## 3.3 Summary of Benefits

### 3.3.1 The Research Libraries Group

1. This project phase will coordinate the efforts of four large research libraries working toward the long-range goal of a common bibliographic network using a union catalog. By sharing the adequacies and inadequacies of a limited operational system, this goal can be evaluated jointly. From actual experience, the cost models presented above can be verified and planning will proceed with fewer unknowns.

2. If on-line access to the Library of Congress proves to be the optimal approach, the Research Libraries Group can base its systems development on the use of the full LC MARC data base without the concomitant expense of storing the unused records and associated indices and of expending the machine resources in updating these files weekly with the new LC MARC distribution records.

3. The Research Libraries Group will have developed its first cataloging service for its member libraries. This service will offer more prompt access to Library of Congress source cataloging than would be possible by using manual techniques or by maintaining a MARC source file locally. It can eventually be enhanced by providing access to the Library of Congress process information file which will allow libraries not only to acquire cataloging data but also to know which titles LC has acquired and intends to catalog. RLG libraries can expect this service to reduce cataloging and processing costs compared to manual alternatives.

4. By having RLG libraries begin to catalog a portion of their materials on a single system, a first step has been made for the development of a common RLG union catalog. Cataloging products for maintaining present card and book catalogs are provided while retention of the machine-readable cataloging record will eventually provide the opportunity to supersede and improve on these two currently used catalog formats. Bibliographic services to library users will not have been interrupted while planning for an on-line union catalog proceeds.

5. This project phase will provide the Research Libraries Group management and technical personnel valuable experience in developing and maintaining a reliable on-line bibliographic system. In addition, the project will provide the opportunity to set up, use, and evaluate accounting procedures related to fee-for-service operations.

### 3.3.2 The Library of Congress

1. The goal of a National Bibliographic Service is ambitious and complex. This project will launch the Library of Congress on an evolutionary development which will reaffirm its role as the national library by using new technologies to serve the bibliographic needs of the nation's libraries.

2. The Library of Congress will obtain actual operational experience in providing on-line services to a network outside the Library. The reliability and responsiveness of this service using current hardware and software will be evaluated. This experience will provide valuable empirical data regarding the future requirements for the National Bibliographic Service. The additional processing loads on the Library's computer equipment will be monitored and used to project equipment needs when several networks would be linked to the service. Data to be collected and presented as statistical summaries are listed in section 3.4.3. Study of the service costs will define the pricing constraints under which this service can be offered by the Library of Congress and yet still be beneficial to libraries.

3. Operational experience will provide information for the development of standards which do not yet exist. For example: (1) definition of a "transmission line" MARC format if the existing format is found to be too inefficient; (2) definition of ancillary physical formats for on-line query and response while negotiating searches for records; (3) content of query and response messages for efficient man-machine interaction.

## 3.4 Technical Approach

This program phase is based on the existence of similar hardware (IBM 370 series computers) and support software (IBM/CICS) at both the Library of Congress and the New York Public Library. Because of this commonality only a modest amount of development work is needed to install a facility in which records are searched for at the Library of Congress, transmitted to New York Public Library, updated on-line, and used to produce catalog card sets. The four major areas of development are communications hardware, communications software, statistical software, and card production software.

3.4.1 Communications Hardware

Both NYPL and LC have terminal control units connected to their main computer to manage terminal traffic for internal on-line applications. Currently, those devices are an IBM 3704 at NYPL and a COMTEN 3670 (which is equivalent to an IBM 270X) at the Library of Congress. Using the necessary line adapters and modems at each end, leased communications lines will be connected to these devices, and the resulting traffic described to each system as if it were originating from compatible terminals transmitting at from 1200-2400 baud.

To handle the traffic produced by two libraries, four "logical" data lines will be needed. (It is expected that only one physical line will be used between New York and Washington, D.C. because the four lines will be multiplexed into a single line at each end of the link.) Fewer lines than terminals are needed because rudimentary line switching software will be included in the NYPL main computer. However, the need for more than one logical line is a consequence of the expedient development approach taken in this phase whereby messages coming from the NYPL computer are treated no differently from any other terminal serviced by the Library of Congress computer. That is, once an RLG terminal has connected with the Library of Congress, it will tie up that line until its search has been either successfully or unsuccessfully completed.

The rationale for determining that four logical lines are required is developed according to the model found in Appendix C to this proposal. As shown in that appendix, best estimates are that use of the LC facility by two RLG members, assuming each member uses an average of 600 LC MARC records per week, will follow this profile:

| | | |
|---|---|---|
| 1. | Average number of LC MARC records used per week retrieved by LC card number (45%) | 540 |
| 2. | Average number of LC MARC records used per week retrieved by author/title key (55%) | +660 |
| 3. | Total average number of LC MARC records used per week | 1200 |
| 4. | Average number of LC MARC records found on local file per week because another library already retrieved the record from LC (assuming 50% duplication of titles between two libraries). | -300 |
| 5. | Total average number of LC MARC records found on the LC file per week | 900 |
| 6. | Total searches routed to LC but are unsuccessful (assume 25% of the total records used, line 3) | +300 |
| 7. | Total searches routed to LC | 1200 |
| 8. | Average time to negotiate a search (in minutes) | 2 |
| 9. | Time LC system available for searching (hours/day) | 8 |

These parameters have been stated conservatively to allow a margin for acceptable system performance. In Phase One, actual performance will be monitored very carefully. According to the model, the following performance can be expected with four data lines:

| | Two Libraries | Four Libraries |
|---|---|---|
| Probability that a terminal will not have to wait for a free line: | 97.9% | 93.4% |
| Average seconds delay when a wait is required | .82 | 3.06 |

It is interesting to note that service to additional RLG members (i.e. from two to four) can be provided without a corresponding linear increase in searches routed to LC because greater advantage is taken of the rate of title duplication between all participating libraries. Therefore, with four libraries parameter #3 would be 2400 LC MARC records used and parameter #7 would be 1725 searches routed to LC. This relationship explains why the additional members can be added with only a small increase in wait time for connection to a line.

However, a critical parameter in the model is #8, the average time a terminal ties up a line once it is connected to LC. If this were to turn out to be 3 minutes instead of 2 minutes, then the probability of no wait is 92.5% or 78.5% and the average delay is 5.3 or 20.9 seconds, depending on whether two or four libraries are using the system. The performance for four libraries under this set of parameters is obviously unacceptable and improvements would have to be made to the communications by adding another line. Another important element is the reliability of the Library of Congress system which must let us spread our service load over an eight hour period each day.

We well understand the limitations to the communications approach taken in this project phase:

- There is no message switching which would allow messages destined for LC to be routed directly without having to be processed in the NYPL computer.
- There is no bi-directional query/response facility to allow LC to access the NYPL computer.
- No mix of different types of terminals can be permitted.
- No cost-effective method of connection by or to other organizations or networks is possible.

All these limitations are to be dealt with concurrently during the development of later program phases. However, despite these limitations which severely inhibit network growth potential, the Phase One approach retains the potential of an attractive and economically beneficial service to the RLG libraries.

3.4.2 Communications Software

This area of development includes all computer programs which support search of LC files and transmission of full records back to NYPL. It will include software changes both at NYPL and LC.

A query for a bibliographic record which cannot be satisfied by files held on-line by the NYPL system will result in a prompt to the user to direct the search to the Library of Congress. The user will at this time rekey the request in conformity with the search requirements of the LC system, as the NYPL system will not attempt to automatically reformat the original request directed to NYPL files. This requirement for rekeying will simplify the software development which at present cannot be directed toward a standard search key protocol. Until such a standard is developed, it would be undesirable to create such software. Use of the Library of Congress system is described in Appendices D and E where the types of search keys and the procedures for negotiating a search are covered.

The Library of Congress system will be modified so that it can detect certain terminals that require transmission of the full record. Upon successful termination of a search at these terminals, a record in the MARC II communications format and ASCII character set will be built in the Library of Congress system and delivered to the NYPL computer. Here the MARC record will be decomposed into the screen image already in use by the NYPL system, and it will be written to the terminal. The record will be treated from that point as a new record just as if it had been locally keyed at the terminal. The terminal operator can make whatever modifications are desired and then continue with the normal procedures for entering a new record into the cataloging data base.

Appendix F to this proposal outlines and flow-charts the software logic which will be developed. In summary, the software modules to be developed during this project phase are:

3.4.2.1 Software Installed on NYPL computer:

1. Accept LC Search Key. This module is invoked by a transaction code from a terminal which indicates that the LC data base is to be searched. The search key presented in the transaction is edited for validity. Then the module "connects" the transaction to one of the four communications lines linked to the Library of Congress. Once the transaction is connected, this module stores the transaction in the CICS transient area, which has the effect of invoking another module to continue processing. If transactions must be queued for a line, this module will continue to attempt to connect at some regular inteval, and if some time threshold is passed without a transaction being connected, the module will notify the terminal operator that the system is busy.

2. Communicate with LC. This module transmits messages to the Library of Congress on the line selected by the previous module. It controls the use of the line by releasing it for other messages if there is no expected reply for the current message. When a reply is received, the module formats the data as a screen image, if necessary, and stores it in the transient area for routing to the appropriate terminal by another module. If the reply is the full MARC record, certain error checking may have to be performed on the record depending on the error checking facilities available in the computer system operating software. Retransmission requests will be sent to LC for MARC records found to be in error.

3. Communicate with Local Terminal. This module transmits and receives local terminal messages while the LC search is negotiated.

4. Accept/Reject LC Record. When the MARC record is displayed on the terminal screen at the end of a successful search, the operator has the option of not using the record at all (perhaps it does not match the title to be cataloged), of modifying the record for local use, or of using the record without change. In the first case, the record is rejected and deleted. In the latter two cases, the operator accepts the record but may have introduced changes on the screen image. The accepted record sent back from the screen is compared with its original form as received from LC. If the record has no modification except call number, only a single copy of the record is stored in the data base. If the record is modified more extensively than just call number, the original LC version of the record is stored in the data base (for use by another library) and the locally modified version is also stored with an indicator that it is a locally modified version of an LC record.

3.4.2.2 Software Installed on LC Computer

1. Format and Send LC MARC II Record. This module is actually a modification of the current LC system which will make the system sensitive to the need of certain terminals to have the full record transmitted in the MARC II communications format rather than as a screen image. Code must be developed to build this record.

2. Retransmit LC MARC II Record. The record sent in the communications format is saved in case a transmission error is detected upon receipt. A new transaction response at LC will retransmit the record.

3. Accept Acknowledgement. When a record has been sent and determined to be free of error, a transaction will be sent to LC acknowledging a valid transmission. LC will then release the record and any temporary storage which may have been saved.

4. End Transaction. This module will respond to a transaction which signals LC that an NYPL terminal is finished even if in the middle of a search. The transaction will invoke normal LC end processing to close in an orderly way.

### 3.4.3 Statistical Software

As noted earlier in this proposal a major benefit of Phase One will be the ability to gather data on the behavior of the system. This information will be crucial to future network development for both RLG, the Library of Congress, and the library community in general. Much will have specific application to the software design of the mini-computer communications controller to be developed during later phases of this program. Phase One software will include the ability to log information about transactions as they occur and then later to organize this data into a useful format.

In all cases where the concept is applicable, we shall attempt to derive frequency distributions rather than simply means, modes, associated standard deviations, etc. This is because distributions are more descriptive of system performance and because we can always derive gross statistical averages from them.

"Message types", as used below, unless otherwise qualified, will mean the three major categories of data transmitted and received by the system: (1) command, (2) display text, and (3) machine-readable data bound for further processing in the receiving machine, e.g. a MARC record. The statistics to be presented are:

1. Distribution of message length, by type of message, and a cumulative distribution including all message types.

2. Distribution of interarrival times for messages, by message type and cumulated. The result would be a plot of time along the horizontal axis and would show for any given time interval the probability that a second message will have followed the first as a displacement along the vertical axis.

Both (1) and (2) should be repeated to show the distribution of inbound and outbound traffic at the NYPL computer. They can then be reversed to give information regarding the situation as observed by the LC computer.

3. Distribution of duration of a logical transaction period. This should be presented in at least three ways: (1) For all transactions, (2) for those that result in success, and (3) for those that result in failure to access a desired record.

4. Distribution of the number of transmissions required to complete a logical transaction. Presented as in (3) above.

5. Distribution of system load by time of day, i.e., logical transactions/unit time plotted against time of day.

6. Ratio of transmission errors to total traffic.

7. Distribution of systems responses over defined states, e.g., relative number of responses which return a MARC record, those which find the record already in the local file, those which result in no record held by LC, those which result in the inability to continue the transaction because of system errors that result in termination of the transaction, etc.

8. Distribution of interarrival time between logical transactions.

9. Distribution of waiting time to find a free line.

10. Distribution of active lines; this would show the fraction of time in which no line was active, one line only was active and all other free, two lines only, etc.

11. Distribution of response time by transaction type.

12. Rate of addition of new MARC records found at LC per unit of time, to NYPL and CUL data bases.

13. Distribution of queries by query type: request by card number, by title key, by author/title key.

14. Ratio of successful queries to unsuccessful queries by key type.

15. Distribution of return records by record characteristic, e.g., language, place of publication, date of imprint, etc.

As part of this project phase, project staff will organize the material for use for the library technical community by issuing a technical report and submitting summarized results for publication as a journal article.

### 3.4.4 Card Production Software

Columbia University has a card printing system known as the MARCARD utility which prints sets of cards from sequential files of MARC communication format records. Since this program uses the MARC communications format as an input record format, it can be used as a card printing utility in conjunction with any system capable of writing sequential files of MARC records. The program has been used to print monographic card sets directly from weekly tapes of the MARC Distribution Service, and to print map and manuscript collection card sets from MARC records entered locally at Columbia University. Since this program operates independently, it is suited to batched printing of cards overnight from files of MARC records that would be selected by an on-line system dedicated to MARC record searching and record entry during the daytime.

The EXPLODER, a key program in the MARCARD system, is programmed in SPITBOL, a high level text editing language developed by Bell Telephone Laboratories and the Illinois Institute of Technology. The language itself absorbs most of the more troublesome aspects of programming for variable length text characteristic of bibliographic data. A general design document for the system is included as Appendix G. What follows is a general description of current capabilities and capabilities to be added under this proposal.

3.4.4.1 Present Capabilities

The EXPLODER program in its present form provides numerous options including:

- provision for either ASCII or EBCDIC character sets
- upper case, upper/lower case, and diacritical output
- varying card stock size

In addition, a set of tables controls the program logic to provide customized card sets to meet a variety of specific requirements. These tables are:

1. Collections. A Collection Table specifies collection identifiers to optionally print at the tops of cards, the classification from the MARC record to be used to direct the cards to a library collection, and the catalogs within the collection for which entries are to be printed.

2. Catalogs. A Catalog Table specifies the catalog indentifier and the set of cards to be printed for each catalog. For each catalog, different card sets may be specified to accommodate union catalogs, departmental catalogs, divided catalogs, and variations for new or added copies. The Collection Table mentioned in the preceding paragraph may also specify card sets for a catalog, which will over-ride the Catalog Table for further flexibility.

3. Card Sets. An Entry Table specifies the types of entries for the card sets to be printed for each catalog. A card set may be defined at any level of MARC encoding, as for instance by form of entry (e.g. conference names) or by function of entry (e.g. subjects) or any combination. Together, the Collection, Catalog and Entry Tables may be encoded to distribute cards over any number of catalogs by any selection criteria.

4. Card Format. The cardform table specifies the formats of individual cards and permits variations in paragraphing, spacing, indention, order and elements to be included.

3.4.4.2 Capabilities to be Added

In this project additional development is required to integrate this program into a production system capable of running at a site other than Columbia University and of handling the needs of catalogs in several institutions. The following features must be included:

1. Table Maintenance. For production purposes at other installations, programs are required to facilitate the maintenance of tables, parameters, and divider card addresses as files which do not require compilation each time the system is run.

2. MARCARD Utility. The EXPLODER program presently prints cards one-up directly on a computer printer. In a production environment, its output should be printed two-up to achieve the best economies with card stock and computer print time. Two additional simple programs are required: a Two-Up Collator program which would split the EXPLODER's sorted output file at mid-point and collate the two halves of the file for printing side by side on card stock which would then be cut into two separate sequences; and an alignment program for aligning the two-up card stock on which the collated file is to be printed.

3. Sequence Control. The EXPLODER program presently produces cards in the sequence of the input MARC file. For each MARC record, the internal sequence of card sets is controlled by the program tables. For the production system, it would be necessary to arrange the cards in a sequence useful to the receiving libraries. A SEQUENCE module will be added to the EXPLODER program to build and attach sort keys to entries under the control of the tables. In addition, divider cards for identification will be generated between sequence sets, providing the address of the recipient. Two sequence options will be provided. At the catalog level or higher, card sets may be sequenced by any unique field such as LC Card Number, Main Entry or Title, or by a locally entered field such as accession number of cataloger. The second option, for output segmented by catalog, would provide alphabetized interfiled card sets.

4. Catalog Card Format and Sequence Control. An additional feature desired in the production system is card format variations between catalogs. The CARDFORM table of the EXPLODER program presently allows variations in card formats according to their entry type, but all catalogs in a processing run must then share the format specified for each entry type. By expanding the Catalog Table to include format codes, card format variations between catalogs can be introduced. This feature is essential to provide a service to not only RLG member libraries but also to components of each member library.

5. Mixed Input. Additional modifications to the EXPLODER program are desirable to allow input of mixed MARC formats (e.g., maps, films, serials, etc.) in a single file. Modification is needed for the EXPLODER to recognize record types and to access the appropriate table each time a new type of record is encountered. A related modification is needed for the EXPLODER to recognize ISBD records and to insert punctuation when these are encountered.

6. Call Number Formatting. The EXPLODER presently formats the call number as a single line at the top of the card which permits optimal use of the limited space on a 3x5" card. However, this approach departs from traditional library practice where the call number is displayed in "blocked" form on the left side of the card which corresponds to the format of the call number on the spine of a book. In order to insure that this system can produce cards compatible with current card catalogs within RLG, a capability will be added to provide a variety of call number format options.

## 4. PHASE TWO AND PHASE THREE DEVELOPMENT

### 4.1 General Description

Phase One of this proposal addresses an important set of needs, namely the gathering of operating experience and hard statistical data when one computer routes requests from its group of terminals to another computer and data base, the specific application being access by RLG terminals to bibliographic information using an on-line facility at LC. As noted in the previous section, the Phase One technical approach has severe limitations when applied to any conceptual design for a national library network.

- Appropriate source cataloging data will not always be found by searching both RLG and LC data bases. Much of the data not found at RLG and LC might be available on-line on some other network data base, e.g. Stanford University's BALLOTS or the University of Chicago system.

- To allow future compatibility for possible cooperative cataloging arrangements, the Library of Congress will need the ability to access bibliographic data generated by RLG libraries or libraries associated with some other network.

Phase One is limited to a single network, RLG, acquiring one type of library data, cataloging, from a single source, LC. Data base access and data transfer is possible only in one direction, LC to RLG.

When the Library of Congress and the Research Libraries Group participate in a national network, there will be an evident need to connect to more than the RLG and LC data bases and such connection should be capable of bi-directional data exchange. A national network must achieve a reliable and economic means for the interconnection of several computer systems together so that at minimum:

- Programs in one machine may exchange records with another machine in various formats for various purposes (e.g. author-title searching, subject searching, book ordering, inter-library lending, etc.).

- Data exchange may occur on the network between computers from different manufacturers. Furthermore, the network should be insulated from differences in system software such as operating systems and terminal monitor systems.

- Mainframe computer failure at one point in the network should not preclude interaction between all other facilities.

- A network facility should not require more than two communications lines in order to have access to network facilities. Computers added elsewhere on the network do not require more lines at all other sites.

In addition to these network requirements, there are other desirable capabilities needed to reduce the communications processing load in the mainframe computer and to insulate complex bibliographic applications programs from contingencies of communications equipment.

- Application programs should not be terminal sensitive, such that a change in terminal type requires maintenance for the application program.

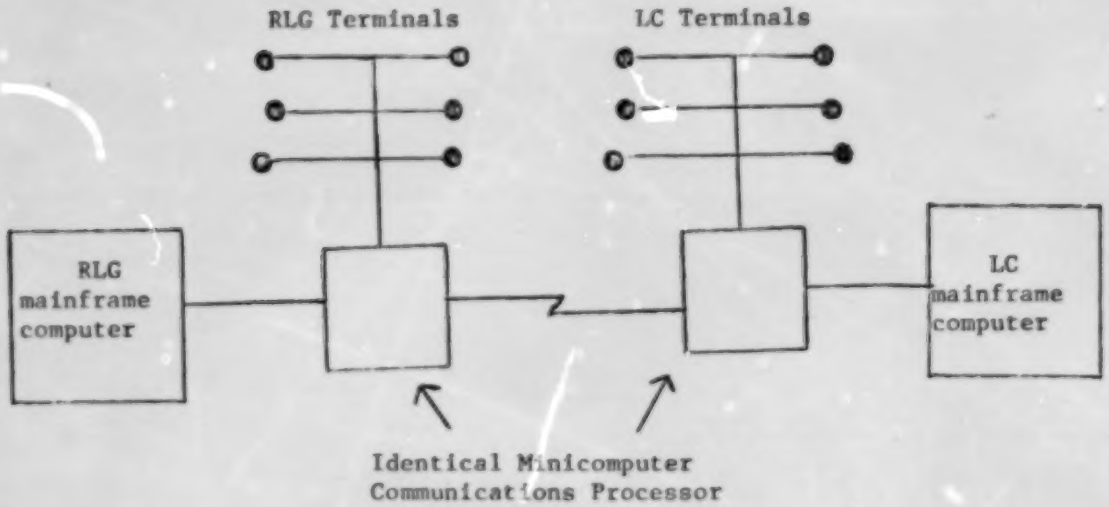
- A transaction originating on a terminal in one facility and destined for another facility should not have to be routed through the local mainframe simply to access data elsewhere. For example, a terminal transaction requesting data from the Library of Congress should not tie up expensive mainframe computer resources locally merely to forward the transaction to Washington, D.C.

If a communications facility meeting these requirements were developed, many of the technological conditions for establishing a national library network would exist. No one in the library community is known to be actively addressing this need for advanced telecommunications capability for libraries. Yet such capability must be available if the Library of Congress is to launch the National Bibliographic Service alluded to in section 2.2 and if the nation's libraries are to benefit from this new service.

The concept of interlinking a number of host processors by use of minicomputer based communication processors is not new. Indeed the technical literature is replete with papers addressing various aspects of such systems (e.g. optimal network topologies, traffic routing strategies, fault detection and techniques for dynamic network reconfiguration to bypass faults, recovery strategies for garbled messages, optimal types and sizes for message blocks, use of various types of processor to processor data links of a variety of transmission characteristics and data rates, detection and correction of traffic bottlenecks, techniques for monitoring performance, etc.). Undoubtedly, the most widely known such network in existence is ARPANET, which was designed and implemented by the firm of Bolt, Baranek, and Newman (BBN) under contract to the Advanced Research Projects Agency. Recently BBN created a subsidiary, Telenet Corp., which has filed a tariff with the Federal Communication Commission to operate a data transmission common carrier network. For some years, Tymshare, Inc. has operated a service (TYMNET) in connection with its other business under a little used provision of FCC Tariff 260 (a tariff filed by the American Telephone and Telegraph Company). ARPANET has connected the mainframes of several defense contractors, TELENET has just begun public service, and TYMNET has interconnected public users for a number of years. Perhaps the user of TYMNET best known in the library community is the National Library of Medicine, which uses that network to make available such services as MEDLINE, SDILINE, etc.

To date all such networks operate under a constraint: because they are common carriers, they may not perform any application-oriented processing based on message content. This prohibition would introduce significant shortcomings for use by a group of on-line information storage and retrieval systems. A network for the national library community needs to be more than one that "passes bits" from one place to another. What is needed is a network made up of a series of interconnected intelligent communication processors (i.e. containing some standardized decision-making programming) which can be adapted to a variety of common carrier services.

Under Phase Two and Phase Three proposals, such an advanced facility will be developed through the use of a minicomputer communications processor using so far as possible existing technology. Two identical minicomputers will be acquired and identical hardware and software components will be installed on these devices to control communications traffic for the Research Libraries Group and the Library of Congress. At this point the incipient network will consist of two facilities capable of bi-directional data exchange in a configuration which schematically can be represented as:




As part of the design effort for this project phase, technical staff from other existing or emerging library network facilities will be invited to participate as consultants. One important activity for this group will be to advise on the selection of a protocol by which the minicomputer processors will communicate. It is anticipated that one of the several new synchronous data transmission protocol standards will be chosen. Final agreement on an industry protocol standard has not yet been reached at national and international levels. However, following the recommendation of our consultants, our selection of a protocol will probably create a de facto library network standard having the force of agreement between library networks and the Library of Congress. This agreement will be reached fully cognizant of the possibility that system adjustments may have to be made when a national standard is set.

Another area of agreement which must be reached will be protocol at the next higher level, message content and format. A communications standard for transmission of the full bibliographic record currently exists, the MARC format, and can be used in an on-line environment until experience shows that it should be changed. Similar standards will have to set for other types of message transmission such that each operates independently from any particular network mainframe computer, processing software, or terminal type.

Having established a network protocol, the simple two facility network illustrated above could soon be expanded either by replication of the minicomputer processor developed under this proposal or by implementing the standard protocols on another minicomputer (which some facilities already own). Existing network relationships need not be altered by the existence of the technology to be developed under this proposal. Direct communication links will continue to be present where the volume of message traffic can make economic use of a line. Likewise, the same principle will apply to the actual configuration of the network. The existence of the minicomputer processor provides the flexibility to allow line connection between network facilities according to the most economic model.

The installation of a national library network will require a very large development effort which will need careful planning and the coordination of many disparate groups. The three phases of this program do not purport to bring into existence that system. Rather they attempt to develop a technical approach on which planning for a national system can be based. Although under this effort the problem of a standard machine-to-machine communications protocol is addressed, there will be many other important standards that will yet have to be developed at a national, if not international, level.

In spite of these limitations, development of communications capability must proceed if libraries are to take full advantage of several services which are now available on remote facilities. Current software and hardware technology reduces the risk of too rapid development by allowing subsets of relatively small developmental scope to be installed, as they are developed, into a properly designed framework. Thus as each component or logical function is completed, that function can be added to a repertoire of functions and made available to a community of users. Phases Two and Three of this development effort offer, therefore, not only a solution to some immediate communications problems but also an approach which by isolating the bibliographic applications on the network mainframe facility can permit flexible adaptation to future advances in telecommunications technology and data standards.

Project Phase Two will include these activities:

Detailed specification of functional requirements;

Standardization of network protocols;

General system design;

Hardware selection;

Project plan for Phase Three;

Project proposal for Phase Three.

Project Phase Three will implement the system specified by Phase Two and will include:

Detailed system design including programming specifications;

Hardware acquisition and installation;

Writing and testing of program code;

Integration and testing of all processor modules in a pilot operation.

## 4.2 Summary of Features

The minicomputer communications processor to be developed under this project phase has the following features:

- Permits access to multiple remote network facilities;
- Allows bi-directional data exchange between network facilities;
- Isolates computer mainframes from communications traffic permitting different types of computers in the network;
- Standardizes network communications protocol;
- Permits flexible configuration of network to adapt to changing traffic patterns and telecommunications economics;
- Provides for efficient message concentration and message switching without using mainframe computer resources;
- Isolates bibliographic applications program residing in mainframe computer from different terminal types;
- Provides a software structure which anticipates future enhancements as new telecommunications technology and standards become available.

## 5. SUMMARY OF PHASE ONE PROJECT PRODUCTS

### 5.1 Computer Software

A set of small logic modules to run on IBM 370 equipment under the CICS terminal control monitor:

Accept LC Search Key;

Communicate with LC;

Communicate with local terminal;

Accept/reject LC record;

Format and Send LC MARC II record;

Retransmit LC MARC II record;

Accept acknowledgement;

End transaction.

In addition, the above modules will log information about transactions as they occur. Appropriate software will be adapted or developed to produce the statistical analyses called for by the project.

All software will be accompanied by appropriate technical documentation.

### 5.2 User Documentation

The following two documents will be prepared by development staff and will be ready for use at the beginning of the operational phase of the project:

A general description of the system from the user's point of view; i.e., the system configuration, processes supported, and products produced;

A detailed manual for terminal operators explaining how searches are conducted on the device, how records are updated, and how output is specified. The manual will be terminal-specific by indicating the function of the various keys present on the keyboard.

### 5.3 Reports and Publications

A series of written reports and articles will be produced by project staff as follows, in the order they are most likely to appear:

News release to library press announcing start of project and its purposes;

Project progress reports prepared according to frequency and detail required by funding agency;

Project final report prepared at the end of the operational phase. This report will be jointly authored by the Library of Congress and the Research Libraries Group staffs. The purpose of the report will be to describe the use of the LC facility for MARC access by the research libraries participating in the project and to make recommendations on the configuration of future RLG and LC systems based on the findings of the project;

A journal article based on the final report if the latter is not in a form suitable for publication;

A technical report suitable for publication which describes the behavior of the on-line system, using the statistical data gathered during the project's operational phase.

## 6. PROJECT PLAN

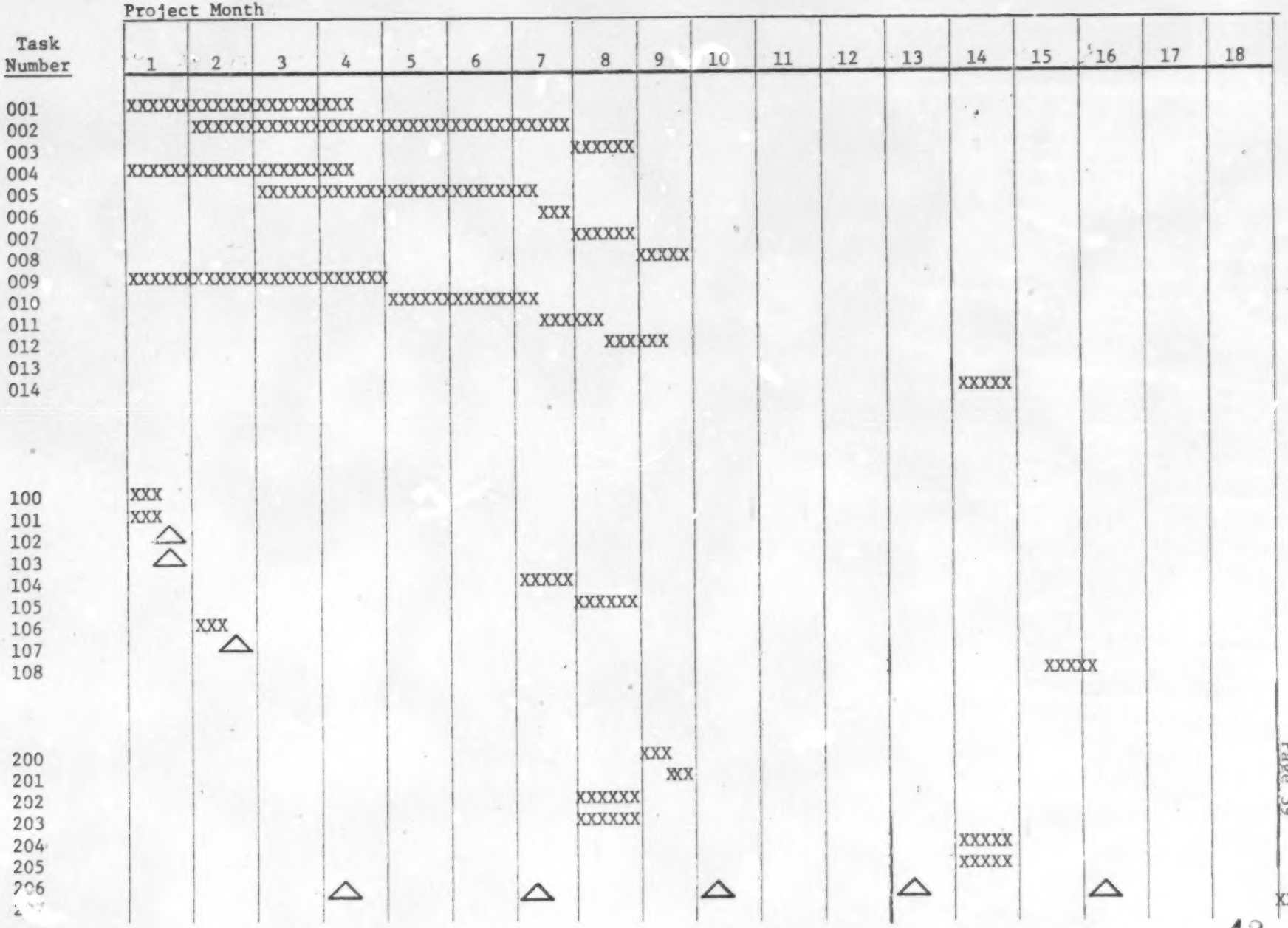
### PROJECT TASK LIST

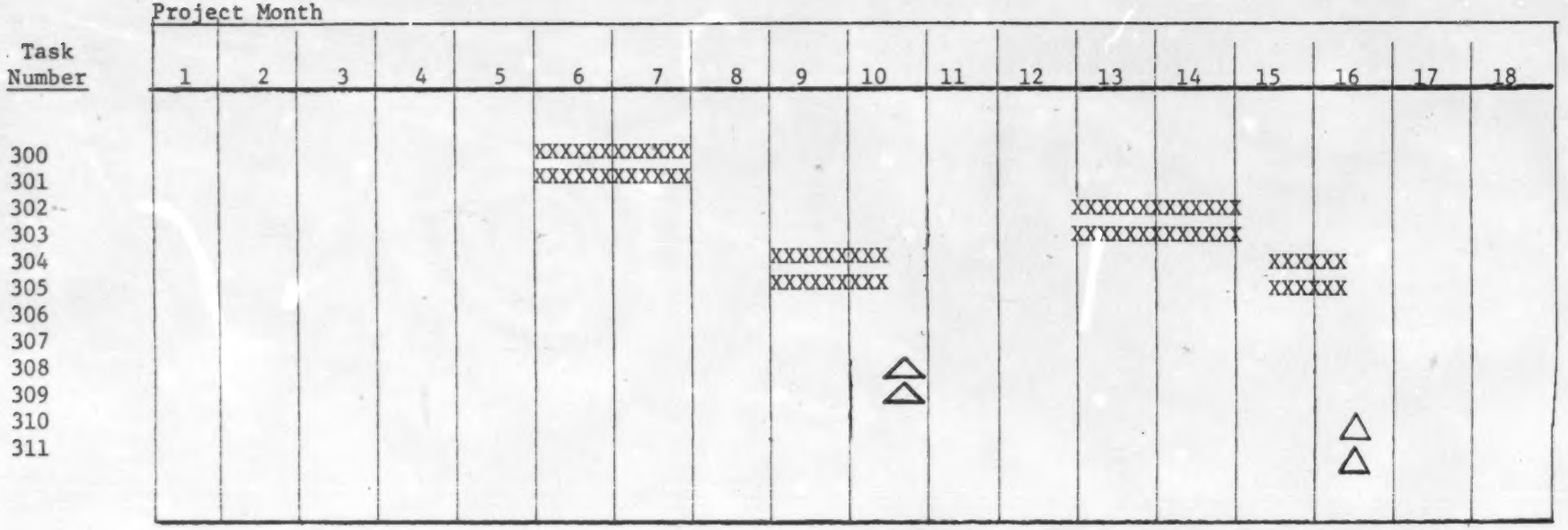
| Number | Task or Event | Development Staff Effort Estimate (man weeks) |
|---|---|---|
| | **Software Development** | |
| 001 | Design programs executed @ NYPL | 10 |
| 002 | Code programs executed @ NYPL | 20 |
| 003 | Test programs executed @ NYPL | 3 |
| 004 | Design programs executed @ LC | 4 |
| 005 | Code programs executed @ LC | 8 |
| 006 | Install programs @ LC | 1 |
| 007 | Test programs @ LC | 3 |
| 008 | Test on-line system | 6 |
| 009 | Design card production system | 15 |
| 010 | Code card production system | 8 |
| 011 | Test card production system | 2 |
| 012 | Install card production system @ NYPL | 2 |
| 013 | Develop, test, install card production tables for Harvard | 3 |
| 014 | Develop, test, install card production tables for Yale | 3 |
| | **Communications** | |
| 100 | Configure communications link NYPL/LC | .5 |
| 101 | Configure communications link NYPL/Columbia | .5 |
| 102 | Order communications gear | -- |
| 103 | Order terminals | -- |
| 104 | Install and test communications link NYPL/LC | 2 |
| 105 | Install and test communications link NYPL/Columbia | 2 |
| 106 | Configure communications link NYPL/Harvard | .5 |
| 107 | Order communications gear | -- |
| 108 | Install and test communication link NYPL/Harvard | 2 |
| | **Documentation** | |
| 200 | System users overview | 1 |
| 201 | CRT users manual | 2 |
| 202 | Procedures NYPL | 2 |
| 203 | Procedures Columbia | 2 |
| 204 | Procedures Yale | 2 |
| 205 | Procedures Harvard | 2 |
| 206 | Project progress reports | -- |
| 207 | Project final report | 4 |
| | **Operations** | |
| 300 | Installation planning - NYPL | 2 |
| 301 | Installation planning - Columbia | 2 |
| 302 | Installation planning - Yale | 2 |
| 303 | Installation planning - Harvard | 2 |

| Number | Task or Event | Development staff Effort Estimate (man weeks) |
|---|---|---|
| | Operations (Cont'd) | |
| 304 | Staff training - NYPL | 1 |
| 305 | Staff training - Columbia | 1 |
| 306 | Staff training - Yale | 1 |
| 307 | Staff training - Harvard | 1 |
| 308 | NYPL Operational | -- |
| 309 | Columbia Operational | -- |
| 310 | Yale Operational | -- |
| 311 | Harvard Operational | -- |

## PROJECT CALENDAR - PHASE ONE

**Project start date:** April 15, 1976




Project Month
Task Number
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
300
301
302
303
304
305
306
307
308
309
310
311

## 7. PHASE ONE PROJECT PERSONNEL

The attached vitae* describe individuals who will work on the Phase One project. Some are directly supported by project funds; others are not. The following lists project personnel and summarizes their role in Phase One:

| Research Libraries Group | Role in Phase One |
|---|---|
| James E. Skipper | Principal Investigator |
| John F. Knapp | Project Manager |
| Peter S. Graham | Assisting Mr. Knapp in project control |
| **Columbia University Libraries** | |
| Jerome H. Yavarkovsky | Manager of Planning and Development at Columbia |
| Heike Kordish | Systems Planning Coordinator for Phase One installation in Columbia cataloging department |
| John E. Leide | Systems Development Coordinator for card system design and installation |
| Thomas Parr | Senior Analyst and Programmer |
| **New York Public Library** | |
| James A. Rizzolo | Manager of Development and Operations at NYPL |
| S. Michael Malinconico | Systems Development Coordinator for systems design and Phase One installation at NYPL |
| Elizabeth A. Romeo | Manager of Systems Analysis |
| Walter Grutchfield | Manager of Programming |
| James G. Hale, Jr. | Systems Programmer |
| **Library of Congress** | |
| Henriette D. Avram | Chief, MARC Development Office |

* Vitae not included in this distribution.

ESTABLISHMENT OF LIBRARY NETWORK ON-LINE ACCESS TO
THE LIBRARY OF CONGRESS MARC DATA BASE

PROJECT BUDGET

18 months, April 15, 1976 - Oct. 15, 1977

DEVELOPMENT - 9 months, April 15, 1976 - Dec. 31, 1976

| | | |
|---|---|---|
| Personnel (including fringe benefits) | | |
| Principal Investigator | $ 3,450 | |
| Project Manager | 7,245 | |
| Development Staff at New York Public Library | 33,401 | |
| Development Staff at Library of Congress | 10,517 | |
| Development Staff at Columbia University Library | 12,440 | |
| Total personnel (Development) | | $ 67,053 |
| Equipment | | |
| Computer Time | | |
| New York Public Library | 6,100 | |
| Library of Congress | 2,300 | |
| Columbia University Library | 12,375 | |
| | 20,775 | |
| Communications | | |
| 2 high speed modem/multiplexors | 2,988 | |
| Installation | 210 | |
| 5 low speed modems | 2,250 | |
| Installation | 344 | |
| Communication line, NYPL to LC | 2,220 | |
| Installation | 138 | |
| Communication line, Columbia to NYPL | 1,056 | |
| Installation | 346 | |
| | 9,552 | |
| Terminals | | |
| 4 Video Terminals | 16,000 | |
| Maintenance | 1,440 | |
| 2 Terminal Printers | 8,000 | |
| Maintenance | 720 | |
| | 26,160 | |
| Total Equipment (Development) | | $ 56,487 |
| Miscellaneous | | |
| Travel | 1,500 | |
| Editorial and reproduction for project reports and documentation | 1,000 | |
| Total Miscellaneous (Development) | | $ 2,500 |
| TOTAL DEVELOPMENT | | $126,040 |

OPERATIONS - 9 months, Jan. 1, 1977 - Oct. 15, 1977

| | | | |
|---|---|---|---|
| Personnel (including fringe benefits) | | | |
| Principal Investigator | $ 3,450 | | |
| Project Manager | 7,245 | | |
| Operations Staff at NYPL | 12,860 | | |
| Operations Staff at LC | 10,517 | | |
| Total Personnel (Operations) | | $ 34,072 | |
| Equipment | | | |
| Computer Time | | | |
| New York Public Library | 60,000 | | |
| Library of Congress | 40,500 | | |
| | 100,500 | | |
| Communications | | | |
| 2 high speed modem/multiplexors | 4,482 | | |
| 9 low speed modems | 4,275 | | |
| Installation (4) | 275 | | |
| Communications line, NYPL to LC | 3,330 | | |
| Communications line, Columbia to NYPL | 1,584 | | |
| Communications line, Harvard, Yale to NYPL | 1,395 | | |
| Installation | 275 | | |
| | 15,616 | | |
| Terminals | | | |
| 20 Video Terminals | 80,000 | | |
| Maintenance | 8,640 | | |
| 6 Terminal Printers | 24,000 | | |
| Maintenance | 2,880 | | |
| | $115,520 | | |
| Total Equipment (Operations) | | $231,636 | |
| Miscellaneous | | | |
| Travel | 1,200 | | |
| Editorial and Reproduction for project reports and documentation | 1,500 | | |
| Total Miscellaneous (Operations) | | $ 2,700 | |
| TOTAL OPERATIONS | | $268,408 | |
| TOTAL PROJECT | | | $394,448 |